

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2001, 00:18:23 ; Search time 28.04 Seconds  
(without alignments)  
924.800 Million cell updates/sec

Title: US-09-494-297-2  
Sequence: 1 MKKTRFKNKLTNTQRYLS.....IAGISLGWGIHTIRKHD 757

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	148	3.8	1183	1	CNA_STAU
2	139	3.5	621	1	HTPG_HELPY
3	139	3.5	1433	1	SUBP_BACSU
4	139	3.5	2366	1	TOXB_CLODI
5	137	3.5	1177	1	Y307_MYCGE
6	136.5	3.5	4590	1	FATH_HUMAN
7	135	3.4	1176	1	SLAP_BACSH
8	135	3.4	1301	1	SLAP_BACST
9	135	3.4	1301	1	SAC3_YEAST
10	132.5	3.4	1116	1	SLPH_BACBR
11	132.5	3.4	3097	1	CADN_DROME
12	130.5	3.3	881	1	LHSL_YEAST
13	130.5	3.3	908	1	DPO1_BORBU
14	130	3.3	621	1	HTPG_HELPY
15	130	3.3	1372	1	FUS1_SCHPO
16	129	3.3	461	1	P55G_HUMAN
17	129	3.3	1292	1	RPOC_MYCGE
18	128.5	3.3	914	1	PBPB_BACSU
19	128.5	3.3	1828	1	MAP2_MOUSE
20	128	3.2	1375	1	GTFC_SPRMU
21	128	3.2	1394	1	HAP_HAEIN
22	127	3.2	984	1	HISA_STRAG
23	127	3.2	1651	1	VIT6_CAEEL
24	126.5	3.2	822	1	DEXT_STRSL
25	126.5	3.2	1772	1	MSPI_PLAYO
26	126	3.2	1036	1	Y414_MYCGE
27	126	3.2	1409	1	HAP1_HAEIN
28	125	3.2	700	1	HS9C_DICDI
29	125	3.2	968	1	CC28_SCHPO
30	124.5	3.2	461	1	P55G_BOVIN
31	124.5	3.2	1251	1	RBP2_PLAYB
32	124.5	3.2	6359	1	BACC_BACLI
33	124	3.1	2334	1	WAPA_BACSU

34	123.5	3.1	1053	1	SLPW_BACBR
35	123.5	3.1	1358	1	SIRA_YEAST
36	123.5	3.1	3063	1	CALC_HUMAN
37	123	3.1	461	1	P55G_MOUSE
38	123	3.1	831	1	LOH_HELPY
39	123	3.1	980	1	BOB1_YEAST
40	122.5	3.1	681	1	BRC2_HALRO
41	122	3.1	663	1	PARC_STAU
42	122	3.1	702	1	ADA3_YEAST
43	122	3.1	1005	1	Y321_MYCPN
44	122	3.1	1608	1	HLVA_SERMA
45	121.5	3.1	1018	1	FNBA_STAU

#### ALIGNMENTS

RESULT 1  
CNA\_STAU STANDARD; PRT; 1183 AA.  
ID CNA\_STAU  
AC Q53654;  
DT 15-DEC-1998 (Rel. 37, Created)  
DR 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE COLLAGEN ADHESIN PRECURSOR.  
GN CNA.  
OS Staphylococcus aureus.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FDA 574;  
RX MEDLINE=92165839; PubMed=1311320;  
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
Lindberg M., Hoeek M.;  
RT "Molecular characterization and expression of a gene encoding a  
Staphylococcus aureus collagen adhesin."  
RL J. Biol. Chem. 267:4766-4772(1992).  
RN [2]  
RP ERRATUM.  
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,  
Lindberg M., Hoeek M.;  
RL J. Biol. Chem. 269:11672-11672(1994).  
RN [3]  
RP COLLAGEN-BINDING DOMAIN.  
RC STRAIN=FDA 574;  
RX MEDLINE=94032261; PubMed=8218209;  
RA Patti J.M., Boles J.O., Hoeek M.;  
RT "Identification and biochemical characterization of the ligand  
binding domain of the collagen adhesin from Staphylococcus aureus."  
RL Biochemistry 32:11428-11435(1993).  
RN [4]  
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.  
RX MEDLINE=97475225; PubMed=9334749;  
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,  
Moore D., Jin L., Schneider A., Delucas L.J., Hoeek M.,  
Narayana S.V.L.;  
RT "Structure of the collagen-binding domain from Staphylococcus  
aureus adhesin."  
RL Nat. Struct. Biol. 4:833-838(1997).  
CC -!- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO  
COLLAGEN-CONTAINING SUBSTRATA.  
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. CELL WALL.  
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS  
IN THE REGION OF THE MEMBRANE ANCHOR.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch).

```

CC -----
CR EMBL: M81736; AAA20874.1; -.
DR PDB: 1AMX; 24-JUN-98.
DR InterPro: IPR001899; -.
DR Signal: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Signal: Repeat; Transmembrane; Cell wall; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 1183
FT DOMAIN 30 1157
FT TRANSMEM 1158 1157
FT DOMAIN 1178 1183
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT DOMAIN 1151 1156
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT REPEAT B1.
FT REPEAT B2.
FT REPEAT B3.
SQ SEQUENCE 1183 AA; 133066 MW; B6A1CC072E575D76 CRC64;

```

Query Match 3.8%; Score 148; DB 1; Length 1183;  
 Best Local Similarity 19.7%; Pred. No. 0.83;  
 Matches 165; Conservative 123; Mismatches 299; Indels 252; Gaps 44;

```

QY 64 PSSSSRYRY-----GYEYVRGHPYK-----QFVAVHDLRVNLEGSRY-----QVYCF 109
DB 182 PDDTHVHVFILINNEKSYVSKDITKDIQIGQQLDSTLNLNVTGTHSNYSQSAIT 241
QY 110 NKKAFPLIS-----DSYKWKYKKHDIKSTFEDYAMSPRING----- 148
DB 242 DEKKAFF-GSKITVDNTKNTIDVTIPQIGSYNSFSINK-----TKITNEQKFEVNS 295
QY 149 -----DELNOKIRAVYNGHPONANGIMEG-----LEPLNALRVYQEAVMYSDNA 194
DB 296 QAWYOHGKEEYVNGK--SPNHVYHNNANAGIEGYKGLKVLKQKDKR-----A 344
QY 195 PLSNPDESK-RESESNLYSTQSLSMRQALKQLIDPNLAKMPKVPDPFOLSTFESF- 252
DB 345 PIAN--VKRLSKKDGSVVKNQD-----KEIETIDANGIANKALPSDYLKEIEAR 397
QY 253 ---DKGDKY-----NKGYNLISGL-----VPTKPPDPMPMPNPQOTTSV 293
DB 398 PTFPDKDKRYPTMKOTDQGYFTTIENAKAIEKTDVGAQKVBETQKVP-----TI 451
QY 294 LIRKAIAGYS-----KILGATLQLTGDVNSFOARVFSSND--IGERIELSDGT 342
DB 452 VFRLKQODNOMTTVPDKAEIKKLEDDGTIKVWSNL-----PENDKNGRAIK----- 498
QY 343 YLTLELN-----SPAGYSIAE-----PI--TFKYVAGVYIIDGKOLENKEI 385
DB 499 YLVKEVNAAGEDTTPGKYKKENGVLVNTERTETTSISGEKVMQDKNOQKREK- 556
QY 386 VEPYSVEAYNDEEFSVLTTONYAKFY-----AKNKNSSQVYVCFNADLKSPSPSEDG 440
DB 557 ---VSNLLANKEGKVTLDVYSETNMKYEFKDLPKYDEG--KIEIYVTVEDHKYDTTDLN 612
QY 441 GKTWPRDFTTGEVKKYTHIAGRLFKYTVKPRDTPDFLKHKKVLEK----- 488
DB 613 GTTINKYTPGELTSATVTKNMDNNQDGRPTEIKVELYQDGKATGTAIINESNNMTH 672
QY 489 -----GYREKGAIEYSGLTETQLRAATQ-----LAIYVTSDAELDKDKLKY 532
DB 673 TWTGLDEKKAGQOVKVTYBELTKVKGYTHVDNNDGNLIVNKKYPTETTSISGEKVMQ- 731
QY 533 HGFQMDNSTLAVAKLIVEYAADSNPRLDDEFIPNNKQOSL---IGTQWHPREDVD 589
DB 732 ---DKDN-----ODGKRPEKYSVA--LIADEKKYKTLVYSETNMKYE--FED 772
QY 590 IIRMDKKEVIEVTNHLRLKRTGTGLAGDRTQDFHEIE-----LKNKQOELLQTVKT- 643

```

```

DB 773 LPKYDEGKKI-----EYTVT---EDHVKYDTTIDNGITITNKKYPGELTSATVTKNM 820
QY 644 -DKTNE-----FKDGKAT-----INLKHGESLTLOGDPE-----GYSLYKE- 680
DB 821 DDNNQDGRPPEIKVELYQDGKATGTAIINESNNMTHWTGTGLDEKAGQOVKTYBEL 880
QY 681 TDSGKYKVVNSQEVANATVSK-----TGITSDETLAFENKKEPVPTGYDOKI--NG 731
DB 881 TKVKGTTTHVDNNDGNLVTNKKYPTETTSISGEKVMQDKNOQDKRPEKYSVNLANG 939

```

RESULT 2  
 HTPG\_HELPY STANDARD; PRT; 621 AA.  
 ID HTPG\_HELPY  
 AC P56116;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE HEAT SHOCK PROTEIN HTPG (HIGH TEMPERATURE PROTEIN G).  
 GN HTPG OR HP0210.  
 OS Helicobacter pylori (Campylobacter pylori).  
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 CC Helicobacter.  
 OX NCBI\_TaxID=210;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=26695 / ATCC 700392;  
 RX MEDLINE=97394467; PubMed=9252185;  
 RA Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.R.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Weidman J.N., Fujii C., Bowman C., Watthey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen Helicobacter  
 RT pylori."  
 RL Nature 388:539-547 (1997).  
 CC -!- FUNCTION: MOLECULAR CHAPERONE. HAS ATPASE ACTIVITY  
 CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC (BY SIMILARITY).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL collaboration  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 3.5%; Score 139; DB 1; Length 621;  
 Best Local Similarity 21.9%; Pred. No. 1;  
 Matches 130; Conservative 85; Mismatches 227; Indels 152; Gaps 30;

```

QY 201 ESFKRESEN-----LVSTQSLSMRQALKQLIDPNLAKMPKQVDDPFLSIFESDQG- 255
DB 28 ETLFLKELVSNASDALDKLNYLMTDEKLR-----GLNTTPSHLSFSDSKKTLITKDNKI 82
QY 256 --DKYN-----KGYONLISGLVPTKPPDPMPMPNPQOTTSVLLIRKAIAGY 303

```

```

Db 83 GMDKNDLIEHGLTAKSGTKNPLSA-----LSGD-----KKKDSALIGOFVGY 127
Qy 304 SKLEGATLDTLGTGNNVSFQARVFSN-----DIGERIEISDGYTTLTSLNPGY---- 354
Db 128 SAFVASKIIVOTKRVSDAAYAMVSDGKGFSEICVDEQGTETILFLKDESDHPSR 187
Qy 355 -----STAEPTFEVGEAGKYTTLIDG-QIENPNKEIPEYSVAYNDEFE----- 400
Db 188 WEIDSVYKKYSEHPEFY-----FLTYDTYHNEGDMOKETKEEKQIINQASALMKMK 243
Qy 401 SVLTQYAKFYAKKNKNGSSOVVYCFNADLSPDSEDDGKWTPTPTGTVKATYH-- 458
Db 244 SELKDKYKFEYQSFADNNEPLSYHNK-----VESLEYTLFLY 284
Qy 459 ----AGDLK-----YVAKPRDTPD-----TFLHKIKVIEKG-----YREKG 494
Db 285 IPSTAPDMFRVYDKSGVKLYKKVFETDDKELLPSYLFKGVISDPLVNSRE-- 342
Qy 495 QAIYVSLTFEQALRAATQALVYFTDSAEIDKDKLDYHGGDMNDSTLAVAKTIVE-Y 552
Db 343 --IIQONKILANITSASVKI--LSEIERLSKDE-KNYHKFE-----PRGKVLKEGLY 391
Qy 553 AODSNPQLTDLDFEIPNNKYOGLIGTQHPEDLVDIIMEDKREYLPVY-HNLTLRKT 611
Db 392 GDFENKKELELRFYSKD--EKLSLKEKENTL-----KENKSIYLLGENDLILKA 444
Qy 612 VTGAGRTYDFHEIEELKNNKOE--LLSOTVKTOKTLEFRDKATINKHGESLTLOG 669
Db 445 SPL--EKYAKQKGYDVL--LDEIDAFVMPGVNEYDKT--PFKDA-----SHSESLKEIG 494
Qy 670 LPE-----GYSYLVKETDSEGYKVKVNSQEVANATVSKTGISDETLAEFN 716
Db 495 LEEIHDEVKQFDLMKAF--ENLKDEIKGEVLSHLTSAVALIGDQONAMMAN 547

RESULT 3
SUBF_BACSU STANDARD; PRT; 1433 AA.
ID SUBF_BACSU STANDARD; PRT; 1433 AA.
AC P16397;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1996 (Rel. 33, last sequence update)
DT 30-MAY-2000 (Rel. 39, last annotation update)
DE BACILLOPEPTIDASE F PRECURSOR (EC 3.4.21.-) (ESTERASE) (RP-I PROTEASE)
DE (90 KDA SERINE PROTEINASE).
GN BPR OR BPF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 195-222.
RX MEDLINE=90110864; PubMed=2106512;
RA Stoma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Therault K.A.,
RA Pero J.;
RT "Bacillopeptidase F of Bacillus subtilis: purification of the protein
RT and cloning of the gene."
RL J. Bacteriol. 172:1470-1477(1990).
RN [2]
RP REVISIONS.
RC STRAIN=168;
RX MEDLINE=90368623; PubMed=2118514;
RA Stoma A., Rufo G.A. Jr., Rudolph C.F., Sullivan B.J., Therault K.A.,
RA Pero J.;
RL J. Bacteriol. 172:5520-5521(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=90216713; PubMed=2108961;
RA Wu X.-C., Nathoo S., Pang A.S.-H., Carne T., Wang S.-L.;
RT "Cloning, genetic organization, and characterization of a structural
RT gene encoding bacillopeptidase F from Bacillus subtilis."
RL J. Biol. Chem. 265:6845-6850(1990).
RN [4]
RP SEQUENCE OF 1-211 FROM N.A.

```

```

RX MEDLINE=89008108; PubMed=1139638;
RA Baill B., Lowe M., Lutkenhaus J.;
RT "Cloning and characterization of Bacillus subtilis homologs of
RT Escherichia coli cell division genesftsZ andftsA."
RL J. Bacteriol. 170:4855-4864(1988).
RN [5]
RP SEQUENCE OF 1410-1433 FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=9017495; PubMed=2106671;
RA Masuda E.S., Anaguchi H., Sato T., Takeuchi M., Kobayashi Y.;
RT "Nucleotide sequence of the sporulation gene spoIIGA from Bacillus
RT subtilis."
RL Nucleic Acids Res. 18:657-657(1990).
RN [6]
RP SEQUENCE OF 195-219.
RC STRAIN=NA10 16;
RA Kato T., Yamagata Y., Arai T., Ichishima E.;
RT "Purification of a new extracellular 90-kDa serine proteinase with
RT isoelectric point of 3.9 from Bacillus subtilis (natto) and
RT elucidation of its distinct mode of action."
RL Biosci. Biotechnol. Biochem. 56:1166-1168(1992).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29035; AAA62679.1; -
DR EMBL; J05400; AAA83362.1; -
DR EMBL; M22630; AAA22458.1; -
DR EMBL; X17344; CAA35224.1; -
DR EMBL; Z99111; CAB13403.1; -
DR EMBL; Z99112; CAB13404.1; -
DR PIR; A35131; A35131.
DR PIR; A35750; A35750.
DR PIR; A36734; A36734.
DR HSSP; P00782; 2SBI.
DR MEROPS; S08.017; -
DR Subtilisin; Bg10233; bpr.
DR InterPro; IPR000209; -
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; signal.
FT SIGNAL 1 30
FT PROPEP 31 194
FT CHAIN 195 755
FT PROPEP 756 1433
FT ACT_SITE 227 227
FT ACT_SITE 274 274
FT ACT_SITE 452 452
FT CONFLICT 219 219
FT CONFLICT 393 393
FT CONFLICT 829 834
FT CONFLICT 836 841
FT CONFLICT 844 852
FT CONFLICT 853 1433
SQ SEQUENCE 1433 AA; 154577 MM; 98DF6846897807C9 CRC64;

```

```

Query Match 3.5%; Score 139; DB 1; Length 1433;
Best Local Similarity 20.7%; Pred. No. 3.5;
Matches 165; Conservative 101; Mismatches 273; Indels 258; Gaps 43;
Qy 64 PDSSEY-----RWGYE-----SYV-----RGHRYKKQFRVAH-----DLRVNLEGSRS 103

```

```

Db 701 PDGSLFLQFKSWHNLDEDDYGVFVLPGEKMKWEOAGVYNGKTSWTDDEIDLKAYKG 760
QY 104 YQVYCNLKAFLGSSSVYK--WYKKHDCISTKPEDYAMSRTIGDEL----- 151
Db 761 Q-----NIQVFNLSQSESTIAKEGWY-----IDVLSLKSAGKYKKKLGYEKP 806
QY 152 --NOKLRAVYNGHPOAN-----GIMEGLEPLNA-IRVQEAIVYSDNA----- 194
Db 807 SGKQKKKPVKPKAKPSANAVKHONKAIOPVLPLKAQVSVETGKSTYSDSTGYTL 866
QY 195 -----PISNDESFRRSESNLVS-----TSQSLMRQALKO-----LIDPML--ATK 235
Db 867 KHKAGDYTLAAEAYGQSKQKSLKTDQTOANFTLEEMKKSTLKGTYINKTGTGEVYG 926
QY 236 MKQVOPDDPOLSTFESRDKDKYNGKQNLSSGLVPTKPTGDDPPMPNOQTTSVLI 255
Db 927 ASYVVEDAAVAPAMTNDKGEYMLEAYE-----GAVTIKAAAG----- 965
QY 296 RKYAIGDYSKLEGAATLQ-----TGDVNSFOARVSSNDIGERIE 337
Db 966 --YYSSEFVELKGDVTKETKALKPFGVYPEGELAVDOSTAENANSYFA--AGNMWAKMT 1020
QY 338 LSDG-----TYTLELNSPAGYSTAEPITTEKVEAGKY--TIIDKQIENPKE 384
Db 1021 LADGDKKGMLTGGLFRWDEFPDPG-----TEFKVE--YVDAIGKDG-----PGKK 1067
QY 385 IVEPYSEAN-----DEEFSVLTONYAKFY-----YAKKNSSQVYVC 426
Db 1068 IAPFNAEALRNGEMWKRVLDSSKGIWVDFLYIQSKPDPYSPGLAMDETQNSGRNV 1127
QY 427 FNDLKLSPDSEDDG-----KTPMPTFTTEGVK--YTHIAGRLDFKYVAPRDTDP 475
Db 1128 QYIDGKMPQDKADGNMIALVDYEAIVPEITSPTDKSTN--KD--SVYTK--GASAP 1181
QY 476 DTFK-----HIKKVIEKGYRE--KGQALIEYSGITFOLRAAQLAI 515
Db 1182 GTTVHLYNGEKEAGETKKAADGTFHAGIILNKGENSELTATASDNGTTD--ASSPTIV 1237
QY 516 YFTTDSAEILDKALDKYHNGCDMNDSTLVAKLIVEAQSNPOLDDLFIPNNKKYQ 575
Db 1238 TLDOEPRELTLDPKD--GKTKNKETLVYKAV--SDDN--VKVNGKA 1283
QY 576 SLIGTOMHPEDLVADITRMEKKEVIPYTHNLTRKTYTGLAGRTKDFHEI----- 627
Db 1284 TVADSGYSARILLENGNEIKVIATDLAANKTKKTYI-----DYNDKVYISGLIP 1335
QY 628 --ELKNNKQELLSTQVTKDTKTNLEFKDGKAT-----NLKHGESLTLQ--GLPEG 673
Db 1336 GEDKNLK--AGESVYKIAFSSAEDLDATFIRMLPTNARASVONATFELPLREISPGRYEG 1392
QY 674 YSLVKEETSEGGYKVVY 690
Db 1393 YMTATSSIRAKKAKVEY 1409

```

```

RA Barroso L.A., Wang S.Z., Phelps C.J., Johnson J.L., Wilkins T.D.;
RT "Nucleotide sequence of Clostridium difficile toxin B gene.";
RL Nucleic Acids Res. 18:4004-4004(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VPI 10463;
RA von Eichel-Streiber C.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBA databases.
RN [3]
RP SEQUENCE OF 1271-2366 FROM N.A.
RC STRAIN-VPI 10463;
RX MEDLINE=92293124; PubMed=1603068;
RA Eichel-Streiber C., Laufenberg-Feldmann R., Sarlingen S.,
RA Schultze J., Sauerborn M.;
RT "Comparative sequence analysis of the Clostridium difficile toxins A
and B.";
RL Mol. Gen. Genet. 233:260-268(1992).
CC -1- DISEASE: CLOSTRIDIUM DIFFICILE PRODUCES TWO TOXINS, AN
ENTEROTOXIN CALLED A AND CYTOTOXIN B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X53138; CAA37298.1; -
DR EMBL: X92982; CAA63562.1; -
DR EMBL: X60984; CAA43299.1; -
DR PIR: S10317; S10317.
DR InterPro: IPR002479; -
DR Pfam: PF01473; CW-binding_1; 18.
KM Cytoxin; Toxin.
SQ SEQUENCE 2366 AA; 269709 MW; E1024BDB8BA56ADF CRC64;

Query Match 3.5%; Score 139; DB 1; Length 2366;
Best Local Similarity 19.6%; Pred. No. 7.4; 249; Indels 296; Gaps 38;
Matches 154; Conservative 86; Mismatches 296; Gaps 38;

QY 83 PYKKQFVVAH-----DLRVNLKESRSYQVYCFNKAAPFLGSSSVYKMY 127
Db 1633 PYFLKFNLTETNTLVYGNQNMIVPENYDLDSDGSISSVINEFSQKLYLGIDSCVKN-- 1690
QY 128 KKHDCISTKPEDYAMSRTIGDELNOKLRAVYNGHPO--NANGIMEGLEPLNAIRTY 183
Db 1691 -----VVISPNITTDENITTPYETNNTYDEVIVLDANYINEKIN--VNINDLS 1737
QY 184 QEAIVYVYSDNAFISNDESFRRSESNLVSQSLMRQALKOLI--DPNLATKMPKO 239
Db 1738 IRYVW-----SNDGNDP--ILMSTSEENKVSQVKIRFVAVFNDKTLANKLS-- 1781
QY 240 VPDDFOLSFESSEKGD-----KYNKGYONLSG--GLVPTKPPPTGDPMPM 284
Db 1782 -----FNFSKDQDVPVSEIILSFTPSY--YEDGLIGYDLGLV----- 1816
QY 285 PNOQTTSVILIRKAYAGDYSKLEGAATLQTDGNVNSFOARVSSNDIGERIELSDGYT 344
Db 1817 -----SLYNEKFYINNFQMWVSG--LVIYINDSLVYFKRPV--NNLITGEVTVGDOKY 1865
QY 345 LTELNSPAGYSTAEPITTEKVEAGKY--TIIDKQIENPKE 384
Db 1866 FNPINGGAA-SIGETIIDDKNYFNOSGVLTQGVFSTEDGFKFAPANTLDEMLDEGEAID 1924
QY 380 NPKKEIVER--YSVAYNDFEESVLTQVYAKFYA-----RNKNGSSQVYVCFA 429
Db 1925 FTGKLIIDENIYEDNVRGAVEMKEIDGEMH--YFSPETKAKFGLNIGDYKYFYNS 1981
QY 430 D--LKSPPDSEDDGKTMTPTFTGEVYKTHIAGR-----DLFKYTK 469
Db 1982 DGYMKGFEVSINDKHYFDDSGVMKVGYTEIDGKHAFYFAENGEMQIGVNTEDGFRY--- 2038

```

```

QY 470 PROTDPPFLKHKKYIEKGYREKGAIEVSGLEFQLRATQAIYFTDUSAE----- 523
DB 2039 -----FAHNDL-----GNEGEIEISYSGILNPNK-----IYFDSDFAVVGWK 2080
QY 524 -----LDKDKLKYHGFQDMN-----DSTLAVAKLIVE----- 551
DB 2081 DLEDSKYYDEDEAEVYIGLSLINDGOYFENDGIMQVGVFTINDKVFYRSDGIIESG 2140
QY 552 -----YAQDSNPQLTDL-----FTFPNN-----KYQSL----- 577
DB 2141 VQNDNMFYIDNGIYQIGVFTSDGYKFAFANTVNDNIYQGAVEYSGLVRGEDVY 2200
QY 578 -----IGTQHPEDVDIIRMEDEKKEVYPTNHLTKRYTGLAGDRKDFHE----- 626
DB 2201 FGFTYTTETG-----IYDMENESDKYFNPET-----KKACKGILNIDIKRYFDEKGM 2251
QY 627 -----TELKNNKQELLQSVTKTKTNLEFKDKQKATINLK-----HGESLTQ-----GLPEGYS 675
DB 2252 RTGLISEFNNNYF-----NENGEMQGYINIEDKMFYFGEDGVQIGVFNTPDGFK 2303
QY 676 YLVKE 680
DB 2304 YFAHQ 2308

RESULT 5
Y307_MYCGE STANDARD; PRT; 1177 AA.
ID Y307_MYCGE
AC P47549;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL LIPOPROTEIN MG307 PRECURSOR.
GN MG307.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
RN NCBI_TaxID=2097;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RC MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Goearne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischman R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fitchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Uterback T.R., Sauder D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA "The minimal gene complement of Mycoplasma genitalium.";
RT Science 270:397-403(1995).
RL [2]
RN
RP SEQUENCE OF 682-1058 FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RC MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RA "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RT J. Bacteriol. 175:7918-7930(1993).
RL
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U39711; AAC71529.1; -
CC DR EMBL: U01767; AAD10584.1; -

```

```

DR TIGR: MG307;
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Lipoprotein; Membrane; Signal.
FT SIGNAL 1 26
FT CHAIN 27 1177
FT LIPID 27 27
FT SEQUENCE 1177 AA; 131881 MW; 276115B041B75F64 CRC64;
SO
Query Match 3.58; Score 137; DB 1; Length 1177;
Best Local Similarity 20.18; Pred. No. 3.4;
Matches 113; Conservative 89; Mismatches 202; Indels 158; Gaps 28;
QY 251 SEDG-DKYNK-----YQNLGGLVPTKPTPPDPMPNQPOTSVLIRKAIG 301
DB 87 NEDKAIKFLKDTKKNVDQYKNTVNGILSP-----PNRRVFIQDQL- 130
QY 302 DYSLLEGA-TLQLTGDNVNSFOARVSSNDIGERI--ELSDGTTLTELNSP---AGY 354
DB 131 DKSQSESMKSSQQLFNQLISDFTAKLFAKDFLYKPRNGOLSTGPYIYDELSQPEKND 190
QY 355 SIAPITFEVAGKVVYTIIDKQIENPKETVEPYSEAVYNDPEEFVSLTQNYAKFYA 414
DB 191 GFQEP-----RESENDALFAKIQAIQIFILWEYTPDLISQATFYA 233
QY 415 KKNKSSQVYICFNAD-LKSPDSEDDGKTYTPDFT-----TGEVYTHIAGDLERYT 467
DB 234 APQOGLGV---VYMERKLD-----KLTPSYAFPPFKDEIRPONQVGNKRMEOI 281
QY 468 VKPRDTPDPTFLKHKKYIEKGY--REKQAIEX---SGLTET---OLRAA 510
DB 282 IEGKDG-----LNTAKGLEKYYIIDQNGNLIDEPPTLISDNTQIKQIYDSINIVQLEAA 336
QY 511 T-QLAIFYETDSAEIDKDKLYHGFQDMNDSTLAVAKLIVE-----YAQDSN 557
DB 337 NLGASLNKLDLNDQKQPLTEIKELNKNLNTIVESTKIENTKSLTLCENHTDSS 396
QY 558 PQLTDL--DEFTPNNNKYQ-SLIGTQHPEDVDIIRMEDEKKEVYPTNHLTKRYT 614
DB 397 QNNKSLIKDAFISNSSNIGKLAQIHTTSDM-----VST 436
QY 615 LAGDRTFQFH--FELEKNNKQELLQSVTKTKTNLEFKDKQKATIN-----LKHGSL 665
DB 437 KASSITSSYLWDAALPNKNTNGASTVSCANATAYQNTSHSNQKLRLVNRGCV 496
QY 666 TLQGLPEGYSYLVEYETDSE-----GYKVVNSQEVANATVSKTITSDET 710
DB 497 AVIGIDGSYLTNNSSEKTERDIEKQKQFLMWRAPQAKTNT--FKNSLYS-FSPINET 553
QY 711 LA--FENKKEPVPTGYDQKIN 730
DB 554 LKTFEKNQEDIL--VNALIN 572

RESULT 6
PATH_HUMAN STANDARD; PRT; 4590 AA.
ID PATH_HUMAN
AC Q14517;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CADHERIN-RELATED TUMOR SUPPRESSOR HOMOLOG PRECURSOR (FAT PROTEIN
DE HOMOLOG).
DE FAT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RC MEDLINE=96163873; PubMed=8586420;
RA Dunne J., Hanby A.M., Poulson R., Jones T.A., Sheer D., Chin W.G.,

```



```

Db 2953 --NRQY--YFIIIGSDPLGFAVETIQNEMKV-----VKKPLDREK 2990
QY 440 --GKTYTPDFTT-----EVKY-----THIAGDLKRYVKKPRDDPDLFLHIKK 484
Db 2991 RDNVLLTLTDTGTFSSKAIVEKVLADANDNSPVCEKTLSDTI-PEDVLP----- 3040
QY 485 VIEKGYRENGQALIEYSGLFTQARATQALAIYFTDSAE-----LDKDK 528
Db 3041 -----GKLIMQISADADIRSNALETIYLLGSGAEKFKLNPDTGELKTSTPLDREE 3091
QY 529 LKDYH-----GFG-----DMNDST--LAVAKILVEYADSNP-----POL 561
Db 3092 QANYHLLVRAITDGGRCQASIVYTLIEDVDNAPERSADYATVPEPTGILLTRVQA 3151
QY 562 TDIDFLIPNNKKYQSLGT--OWHPEDLVDIRME--DKK-----EVIPTNHLTL 608
Db 3152 TDADAGL--NRKILYSLDSDAGQFSINELSGIQLKPLDRELQAVYTLSLKAVDGLPR 3210
QY 609 RKVVTGLA-----GDRKDFHEIEIKNNKQELLSQT-----VKTDKTNLEF-- 650
Db 3211 RLATGTVIVSVLDINDNPPVFEYREYAGVATSDILVGEVLQVYASRDIEANAEITYS 3270
QY 651 -----KDKKATINLKHGESLTLQGL-----PEGYSYLKEDT-----SEGKVKVNSQEV 695
Db 3271 IISGENGRKFSIDSKGAVYIENLDYESHETVLYVEATDGGTSPSLSDVATVNAVNTDI 3330
QY 696 ANAT-----VSKTGITSDETLAFENKKEPVPTGVQDKINGYLALIVAG 740
Db 3331 NDVTPVFSQDTYTVTVISEDAVL--EQSVITVMADDDADGPNSHIHVSIIIDG 3379

RESULT 7
SLAP_BACSH STANDARD: PRT; 1176 AA.
ID SLAP_BACSH STANDARD: PRT; 1176 AA.
AC P38537;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE SURFACE-LAYER 125 KDA PROTEIN PRECURSOR.
OS Bacillus sphaericus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2362;
RX MEDLINE=89327128; PubMed=2666389;
RA Bowditch R.D., Baumann P., Yousten A.A.;
RT Cloning and sequencing of the gene encoding a 125-kilodalton
RT surface-layer protein from Bacillus sphaericus 2362 and of a related
RT cryptic gene.
RL J. Bacteriol. 171:4178-4184(1989).
CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
CC OF PROTEIN WHICH COAT THE SURFACE OF BACTERIA.
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M28361; AAA50256.1; -
CC PIR: A33856; A33856.
CC InterPro: IPR001119; -
CC Pfam: PFO0395; SLH_3.
CC PROSITE: PS01072; SLH DOMAIN; 2.
CC Signal; Cell wall; S-layer; Repeat.
CC SIGNAL 1 30 POTENTIAL.

```

```

FT CHAIN 31 1176 SURFACE-LAYER 125 KDA PROTEIN.
FT DOMAIN 32 91 SLH 1.
FT DOMAIN 92 151 SLH 2.
FT DOMAIN 152 210 SLH 3.
SQ SEQUENCE 1176 AA; 125225 MW; 1A9A458EF433788C CRC64;

Query Match 3.4%; Score 135; DB 1; Length 1176;
Best Local Similarity 18.5%; Pred. No. 4.4; 256; Indels 238; Gaps 31;
Matches 135; Conservative 101; Mismatches

QY 166 NANGIMEGLPEPL-----NAIRVTOEA-----WVYSPNADISNPDESFKR 205
Db 59 DANGNFPPLKTIISRAEATITFTNALELFAQDVNFQDKADARYT-DATAIATYENGFFEG 117
QY 206 ESESNLVSTQSLSMLRQALKOLIDPNLATKMPKVPDDFQSLFSESEKGD----- 256
Db 118 VSAITEFAPNKQLT-RSEPAKLIVD-----AFELGEGDLSSEFADAST 158
QY 257 --KYNKGYQLLSGGLVPTPTPPGDPMPNPOTTSVLIRKKAIDYKLLGATLQL 314
Db 159 VKPMASYLEIAVANGYIKGSEANGKTNLNPNAPITR--QDFAV-VFSRTLENDATP 213
QY 315 TGDVNSFOARVFSNDIGERIELSDGTITLTLSNPAGYSIAEPIFFKY-----EAGKV 369
Db 214 KVDKIEVVDKATLN-----VTLSDGTRETVLEKRLKLEPNKETEVYFKIKDVEYKAKVT 266
QY 370 YTIIDGKQIENPKKEIPEYSEVANDFEESVLTQONYA-----KEYYAKNKG 420
Db 267 YVYTTATVAVKSVSATNLKEVYVEFGDVTGKETAEADANVALKSGKTIKSVSLAADNKTAT 326
QY 421 SOVYVCFN--ADLKSPPDESDGKTM--TPPTTGEVAYKTHAG-RDLFKTYVPRDT 473
Db 327 VTLLDKLNKKADALISNVKAGDKELNVKRVETAVDNKIPEVTEVKSIGTAKVATLS 386
QY 474 DPDFLKHKKVIEKGYRENGQALIEYSGLFTQARATQALAIYFTDSAELEDKL----- 529
Db 387 EP-----VENLSNTNTLDGKA--YFGNVYMGAGKNTVTLTPTSSALVSGHKLVSG 438
QY 530 -KDYHGFQDN-----DSTLAVAKILVEYADSNPQTLTDLDFLIPNN 571
Db 439 AKDFAGVSLNSHFEKVEVDKEAPVTEATLEVTLETFSE-----IDMDTVASN 492
QY 572 -----NKYQSLI-GTQ-----WHPEDLVDI-----IRMDKK 597
Db 493 VYKSGDSKKEASEFERIADNKYKFEKGESEKTLPTGKVDVYVEDIDKYSDNKIATKTV 552
QY 598 EVIP-----VTHNLTRKTYVGLADRTKDFHEIELEKNNKQELLSQT 640
Db 553 TVPEIDQTRPEYKRYVALDEKTIKTFSTVDGESAIKIG--NYVKKDKDK-----V 604
QY 641 VKTDKTNLEFKDGA-----TINLKHGESLT-LOGLPEGYSYLKEDTSEG 685
Db 605 VASVDKTVVDSKDSVYIIDLSKVSAGENTITIKNVADKALNNMTDYGKFTRSDEG 664
QY 686 -----YKKYVN-----SQEVANATVSKTG- 704
Db 665 PDYEHVINADAKAKVYLKPKDKRMDASLADYSNYLVKINDTLQTLSEDAVATLSVNDAT 724
QY 705 ---ITSDETL 711
Db 725 VVTITFAETI 734

RESULT 8
SLAP_BACST STANDARD: PRT; 1228 AA.
ID SLAP_BACST STANDARD: PRT; 1228 AA.
AC P38825;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1995 (Rel. 32, Last annotation update)
DE S-LAYER PROTEIN PRECURSOR (SURFACE LAYER PROTEIN).
GN SBSA.

```







```

Db 611 FPPSQESKQLPQISQSHHTST---NPLITPOVGDLSQKQOQIKTYTDDGSGPPYEDQSA 667
QY 168 NGIMEGLEPLAIRVTOEAWVWYSDNAPISNPDESEKRESSESNLVSTQSLMKRO-ALKO 226
Db 668 QNSTEASKAHMISTJNSGAY---DEKLSSEQEMRKKEQRREBEETQOLKQKQOENADKO 724
QY 227 LIDPNLATAKPKQVPPDPQSLSTFSESDKGDYKKGYNLISGGLVPTKPPDPPMPN 286
Db 725 VITEQIANDLVKKEVNSVYSIVKREFSEANVRKDFDTWTRELY-----DAFLHER 776
QY 287 -----QOPTSVLIRK-----YVIGDYSKLE-----GATLOLQMDN 318
Db 777 LYLIYDSRAELRNSTLKKKFEKKQOASTSQAKKNRILEEKKREIKLYSHQGVPGFK 836
QY 319 VNSFOARVSSNDIGERIELSDGTYYLTTELNSPAGYSIAPIETFEYKGVYTTIIDGQI 378
Db 837 KSPCLERTPTPKGVNNSFMSSSDKNL--IFSPVNDENFKFATHLTKISKLMRPLEMQSI 894
QY 379 --ENPKET-----VEPYSEAYND-----FEFESVLTQNTAKFYAKNKGSSQ 422
Db 895 YYNLTKKKFPPNSILTPANLFIYAKDWTSLSNRWILSKFNLQIADSKKF-----SNNTISS 950
QY 423 VVYCFNADLSPDSED-----GKWTMPDFTTGEVKY-----THI 458
Db 951 RIICID-DEYEDSFSDQLLINTGYT--NPDLFDEMKLKDGEELIKLITGISLNTNI 1008
QY 459 AGHDLKRYTVKPRDPTDFLKHKKRY--TEKGYREKQAIEXSGLT-----ETOLR 508
Db 1009 CFSLLIITYESANNTLSESTIKHLKLNIRSKYSSVIERIDLMLNLTSESPHCLEDKLS 1068
QY 509 AANQAIYYTDSAEIDKDKLADYHGFQMDNSTLAVAKLIVEYADDSNPQLT-DID-- 565
Db 1069 EISHSVYVYKLEKGYDK-----TLRKRSIAGIHSKSTDLQTKTDIDQK 1113
QY 566 ---FFIPNNKKYOSLIG---TQWHPEDLVDIIMEDEKKEVIRP-----THN-----LTLR 609
Db 1114 MKMKEKKKKKQOQIGERTYTAHLESHD-ASPRSKKAKKLPILSTSHSQKRPPLASR 1172
QY 610 KIVTGLAGDRTKDFHEIELKNNKQELLISQVTKDTKNTLEFKD-----GKATINLKHG 662
Db 1173 LMSGSSSTPRLPSHLAMKRRKNSRVTSLSHTVLPVSPHSNNINPAASFEGNNTTDI--- 1229
QY 663 ESLLTGLPEGYSYLVKEDSEGYKVVNSQEVANATVSTGITSDETLAFENKKEPVY 722
Db 1230 -----OSQOLIENOKSTSVLNNVSERIIGNOEICQTP-----NEVTP 1268
QY 723 T--GVDO 727
Db 1269 VLDGADQ 1275

```

```

CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
CC HEXAGONAL S-LAYER.
CC -1- SIMILARITY: CONTAINS 3 S-LAYER HOMOLOGY (SLH) DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: D90050; BAA14103.1; -
DR PIR: A35129; A35129.
DR InterPro: IPR001119; -
DR Pfam: PF00395; SLH; 2.
DR PROSITE: PS01072; SLH DOMAIN; 2.
KW Signal; Cell wall; S-layer; Repeat.
FT SIGNAL 1 53
FT CHAIN 54 116 SURFACE LAYER PROTEIN.
FT DOMAIN 57 120 SLH 1.
FT DOMAIN 121 171 SLH 2.
FT DOMAIN 172 231 SLH 3.
SQ SEQUENCE 1116 AA; 123397 MW; 86D583D7AC72546F CRC64;

Query Match 3.4%; Score 132.5; DB 1; Length 1116;
Best Local Similarity 18.9%; Pred. No. 5.6;
Matches 162; Conservative 112; Mismatches 266; Indels 315; Gaps 45;

QY 53 LVESSTPNAINPPSSSEY-----RWIGSEYVRG-----HPY 84
Db 143 IVKGFPDKSEFPQNOVYAYAVMYIALIGYEPSPVWNSMISKSELNIAKGINPN 202
QY 85 YKQFRAVA-----HDLRVNLEGSRSYOVYCNLKAAPLGS-----SVYKKWYK 130
Db 203 MOQPAATIFKMLNALRYKLMEOIEY-----GTDIRLANTDETLLTKYIK-- 247
QY 131 DGISTKFEDYAMSPRTIGDEL-----NOKLRAVYNGHPONANG-----IMEG 173
Db 248 --VTVRDMDAHEKGNNSDELPLVTNVPAIGLSLKANEVTLNGKADLDGNSNTYKVAEG 305
QY 174 LEPNLAIRVTOEAWVWYSDNAPISNPDESEKRESSESNLVSTQSLMKROLKQIIPNLA 233
Db 306 INP--NAPDGGKQVQWAKIDDENVI---VMEGSEDEDDVMDRVSALYLKGAFTD-DIV 359
QY 234 TKMPKQVPPDDFOLSIPESEBKDKYNGYONLISGLVPTKPPGPPMPNPQPTTSV 293
Db 360 KDLKSDLDVYKIDMGSE----- 378
QY 294 LIRKVAIGDYSKLLEGATTLQLT--GDNVNSFOARVSSND-IGERTELSDGTYYL----- 345
Db 379 --KSYRLTEDTKI---TYNFTRENDPVDAL-SKIYKNDNTPGVKAVLNDNNEVAYLHI 431
QY 346 ---TELSNPAAGYSIAETITTKVEA-----GKVYTI-IDGKOIEN 380
Db 432 DDQITDKSVKGVYRGSKVSIKIDAKKRTITLNDNSKESLEDDEGDFLVFLDGGPAKL 491
QY 381 PNKEIYEPYGV-FAVNDPEEFVSULTQNVAKFYAKKKNSSOVVYCFNADLSPDSED 439
Db 492 GDLKESDVYSVYAADGDKKYLVPANRNVAE-----GKVEKYVSRN---KTDIRLIV 540
QY 440 GGRY--MTPDFTTGEVYKTYTHIAGRDLEKRY--VKPRD-----TDPDTFLKHKK-- 483
Db 541 GKRTYKVPD-----ASYSENANKDVKVNSDLDLISNLGEEVKILLDPSGRVRIETK 595
QY 484 -----KVIKRG--YRKKGQAIIEYSGLETQLRAATQALAIYTFDSLEDKDKIKDY 532
Db 596 DAIDRRKPLAIITKGAAYNSSKDTYPTVMTQ--KKKTOI-----VSLDQKDIYDR 644
QY 533 HG--FGDMNDSTLAVAKLIVEYAQ-----DSNPQLTDLDF-----FIPN 570
Db 645 YGVNYSKSNDRQAFKDLVELLQPVYKVEDSATDANQYLVLEFNDSKGEVDKAVYLD 704

```



```

FT  DOMAIN  978  1078  CADHERIN 7.
FT  DOMAIN  1087  1183  CADHERIN 8.
FT  DOMAIN  1193  1299  CADHERIN 9.
FT  DOMAIN  1307  1414  CADHERIN 10.
FT  DOMAIN  1423  1514  CADHERIN 11.
FT  DOMAIN  1523  1630  CADHERIN 12.
FT  DOMAIN  1639  1742  CADHERIN 13.
FT  DOMAIN  1749  1861  CADHERIN 14.
FT  DOMAIN  1870  1966  CADHERIN 15.
FT  DOMAIN  1974  2085  CADHERIN 16.
FT  DOMAIN  2346  2377  EGF-LIKE 1.
FT  DOMAIN  2407  2585  LAMININ G-LIKE 1.
FT  DOMAIN  2592  2627  EGF-LIKE 2.
FT  DOMAIN  2869  2902  EGF-LIKE 3.
FT  DOMAIN  2661  2822  LAMININ G-LIKE 2.
FT  DISULFID  2346  2357  POTENTIAL.
FT  DISULFID  2351  2366  POTENTIAL.
FT  DISULFID  2368  2377  POTENTIAL.
FT  DISULFID  2592  2607  POTENTIAL.
FT  DISULFID  2601  2616  POTENTIAL.
FT  DISULFID  2618  2627  POTENTIAL.
FT  DISULFID  2869  2880  POTENTIAL.
FT  DISULFID  2874  2891  POTENTIAL.
FT  CARBOHYD  2893  2902  POTENTIAL.
FT  CARBOHYD  97  97  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  150  150  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  325  325  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  426  426  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  930  930  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  CARBOHYD  1266  1266  N-LINKED (GLCNAC. . .) (POTENTIAL).
FT  VARIANT  1425  1425  E -> K (IN ALLELE CADN-M12; MUSCLE DEFECTS).
FT  CONFLICT  1342  1342  P -> A (IN REF. 1).
FT  CONFLICT  2786  2786  S -> T (IN REF. 1).
SQ  SEQUENCE  3097 AA; 347201 MW; 082242F28DB5CC3 CRC64;

```

Query Match 3.4%; Score 132.5; DB 1; Length 3097;  
 Best Local Similarity 19.8%; Pred. No. 26;  
 Matches 114; Conservative 89; Mismatches 192; Indels 181; Gaps 29;

```

QY  287 QPPTSVLRKKA-IGDYSKLLGATLQLT--GDNNSPQARFSSNDIERIELSDGT 343
DB  719 QPHTVSLVATATEDSGFSSVD-LTTRVDVDMNPKFELPYQAHNVDEDIPLGTSIL 777
QY  344 TLETLNSPAG-----YSIAEPTFVEAGKYVTIIDGKQIENPKETIVPSVEAYNDF 397
DB  778 RYKAMDSGSSNAIEIYVSD--HFAVDSNGI--IVNNKQLADANNAYEFIVTAKDKG 834
QY  398 E--EFSVLTQNTAKFYKANKNGS---SQVYCFENADIKSP-----DSED 439
DB  835 EPPKSGVATVR-----VYTKNKNDSEPKFSQYVTPNVDENAGPNTLVTVASDKGDN 889
QY  440 -----GKTKMTDPF-----TGEVAKYTHLA--GRDLKFTV----- 468
DB  890 VRFEGVGGGTSQFYIEDITGVIRLHNNKISLKDCKYELLVTAMDDGSCCVNGDQTIHT 949
QY  469 -----KPRDTPDTFLKHK-----KVI-----EKGYRKGAIEX 499
DB  950 STAVVVVFTIDVANDKNKPVFEDCSTYYPKVEGAPNGSPYKAVATDEKDV--NGO-VKY 1006
QY  500 SGLTEPQLRAATOLAIFYTDSAELEKDKLKYHG-----FGDMNSTL-AVAK 547
DB  1007 S-IYQOPNCKGTAFV--DEETGEVSTNKVDFREDGDKFVSYTVAKATDQGDPSLEGVCS 1063
QY  548 ILVEYAKDSDNPQLTLDLDFIPNNNKYQSLIGQWHPEDLVDIIRMDKKEVLPVTHNL 607
DB  1064 FYVEITDVNDNPPLFPRQKVV-ENVKODASIG-----NLRVS----- 1101
QY  608 LRKTVTGLADRFKDFEIELEKNNKQELLSQVVKTDKTNLEFKDKATINLKHGESITL 667
DB  1102 -----ASDEAD-----NNGAIYVSLTAPFNPDLDFEYFIOA----- 1133

```

```

QY  668 QGLPEGSVLYKETSDESGRYKVNNSQEVANATYSKQITSDENLAFENKPEVPIGVQD 727
DB  1134 ---ESGIVLYKPKLDRETYLLEMAQDKGYPLPSRTVEVDIDVDRANN-----PPVMDH 1185
QY  728 KINGYLAL-----IVYAGISLGIWGIHTIRIR 754
DB  1186 TVIGPIYKEMNPMYGGKVVYSIKASSGIEGNPTVYFVR 1221

RESULT 12
LHSL YEAST
ID LHSL YEAST STANDARD; PRT; 881 AA.
AC P36016;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HEAT SHOCK PROTEIN 70 HOMOLOG LHSL PRECURSOR.
GN LHSL OR YKL073W OR YKL355.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M., Pohl F.M.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=94378724; PubMed=8091863;
RA Rasmussen S.W.;
RT "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT NUP100 gene, an open reading frame (ORF) possibly representing a
RT nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT addition to seven ORFs with weak or no significant similarity to
RT known proteins."
RL Yeast 10:569-574(1994).
RN [3]
RX CHARACTERIZATION.
RP MEDLINE=96256281; PubMed=8654361;
RA Craven R.A., Egerton M., Stirling C.J.;
RT "A novel Hsp70 of the yeast ER lumen is required for the efficient
RT translocation of a number of protein precursors."
RL EMOB J. 15:2640-2650(1996).
CC -1- FUNCTION: PROBABLY ACTS AS A CHAPERONE INVOLVED IN BOTH
CC POLYPEPTIDE TRANSLOCATION AND SUBSEQUENT FOLDING.
CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN.
CC -1- PTM: N-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on any
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: X75780; CAA53401.1; -
EMBL: Z28073; CAA81910.1; -
DR PIR: S37895; S37895.
DR PIR: S39169; S39169.
DR PIR: S44514; S44514.
DR SGD: S0001556; LHSL.
DR InterPro: IPR000886; -.
DR InterPro: IPR001023; -.
DR PROSITE: PS00014; ER_TARGET; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
DR PROSITE: PS00297; HSP70_1; FALSE NEG.
KW Chaperone; Endoplasmic reticulum; Glycoprotein; signal; ATP-binding.
FT SIGNAL 1 20
FT CHAIN 21 881 HEAT SHOCK PROTEIN 70 HOMOLOG LHSL.

```

FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 527 527 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 844 844 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 878 881 PREVENT SELECTION FROM ER (POTENTIAL).  
 SQ SEQUENCE 881 AA; 99571 MW; ACED092CA34785 CRC64;

Query Match 3.3%; Score 130.5; DB 1; Length 881;  
 Best Local Similarity 18.3%; Pred. No. 5.1;  
 Matches 132; Conservative 119; Mismatches 270; Indels 199; Gaps 32;

52 GVESTPAINPDSSS-----EYRMVGEYSYVRG-----HPYKOF 88  
 241 GIKKSMSTLDPEDDTTOVTEFEBSYGINPHLGAKFTMDIGSLIENKFLTHPAIR-- 298  
 89 RVAHDLRVNLEGSRSYQVYCFMLKAPLGSDSVKKMYKKHGDISTKPEDYAMSPRITG 148  
 299 --TDELHAPKALAKINQAERAKILLSANSEASI-----NIESLINDIDFRTSIR 348  
 149 DELNOKLRAVNTNGHPONANGIMEGLEPLNAIRVQEAIVYSDNAPISNPDESFRERE 208  
 349 QEEFEFI-----ADSLIDIVKPIIND-AVTKQ-FGGYGTMLP-----E 383  
 209 SNLV-----STQSLMRQALKOLIDPNLATKMPKROVDDFOISFESEDDGKYNKGYON 264  
 384 INGVILAGSSSKIPYODOLIKLVSE--KYLRVYNAD-ESAVNCGVMRGIKLSNFR- 438  
 265 LLSGLVPTKPPPTGDDPMP-----PNOPQTSVLRIRKAYIGDSKLEGATLQTLQDN 318  
 439 -----TKPLNVDRSVNTYSFKLSNSELYDVFTRSAAVKNKSLILNT-DS 485  
 319 V-NSFOARVSSNDIGERIELSDGTYTTELNSPAGYSTAEPITFEVGEKYYTIID--- 374  
 486 INNNFTIDLFENGKLEFETITVNSGAIKNSYSSDKSCSSGAAYNTFTPLSSDRFSIOEVNC 545  
 375 -----GROIEPNK-----EIVEYSVEAVNDEFEFVLTQONAKFYAK 415  
 546 ICGSENDIGNSKOIKKKGRSLAFTSEDEVETKRLSPERSRHEHITKLDKODEKROFOE 605  
 416 NKGSSOVVYCFNADLSPDSEDEGKTYMTPDFTGEVYTHIAGDLKRYTVKPRDTP 475  
 606 NNVLESNILYDARNLLMDEVMQNGPKSOVEELS--EMKVVYLDWLEDAF-----DTDP 658  
 476 DFFLKH-----KVYIE-----KYTRKG--QAIYSGLTETOLRAA 510  
 659 EDIVSRIRREIGILKKKLEIYMSAKEPLNSOQKGMLEBHKLLQAIETHKNTVEEFLSO 718  
 511 TOLAIYFFDMSAELDKDKDYHGFGDMNDSTLAVAKLIVEAO--DSNPOLTDIDDFI 568  
 719 FETERPDTIDNREEKKIK-----QPAYYSKALSTWEELTSTRKNSISELEKEL 768  
 569 PNNKQSLIGTQWHEDELDV-----IRMEDEKKEVLPYTHNLTKRYTGLAGDRT 620  
 769 AKN-----LFGEDLR-EHLFEIKQFDWYRTKLEKRLIK-----SGDES 808  
 621 -----KDFP-ELELKNNOELLSTV-----KTDKTNLEFGDKRTI-NLKHGES 664  
 809 RNEIKKLHNRFLRLOKREKRLKLEQEKSHNNTESTVINSADKTTIVNDTTES 868

RESULT 13  
 DPOL\_BORBU STANDARD; PRT; 908 AA.  
 ID DPOL\_BORBU  
 AC 051498;  
 DT 30-MAY-2000 (rel. 39, Created)  
 DT 30-MAY-2000 (rel. 39, Last sequence update)  
 DT 30-MAY-2000 (rel. 39, Last annotation update)  
 DE DNA POLYMERASE I (EC 2.7.7.7) (POL I).  
 GN POLA OR BB0548.

OS Borrelia burgdorferi (lyme disease spirochete).  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.  
 OX NCBI\_TaxID=139;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35210 / B31;  
 RX MEDLINE=98065943; PubMed=9403685;  
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,  
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,  
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,  
 RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,  
 RA van Vugt J., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,  
 RA Uitterback T., Watthey L., McDonald L., Artlach P., Bowman C.,  
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,  
 RA Smith H.O., Venter J.C.;  
 RA "Genomic sequence of a Lyme disease spirochete, Borrelia  
 burgdorferi."  
 RT Nature 390:580-586(1997).  
 RL Nature 390:580-586(1997).  
 CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
 CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =  
 CC N PYROPHOSPHATE + DNA(N).  
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-A FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC EMBL: AE001156; AAC66909.1; -  
 CC TIGR: BB0548; -  
 DR DR  
 DR InterPro: IPR001098; -  
 DR InterPro: IPR002298; -  
 DR InterPro: IPR002421; -  
 DR InterPro: IPR002562; -  
 DR Pfam: PF01612; 3.5 exonuclease; 1.  
 DR Pfam: PF01367; 5.3 exonuclease; 1.  
 DR Pfam: PF00476; DNA\_POL\_A; 1.  
 DR PRINTS: PR00868; DNAPOL.  
 DR PROSITE: PS00447; DNA\_POLYMERASE\_A; 1.  
 DR Transferrase; DNA-directed DNA polymerase; DNA replication; DNA repair;  
 KW Hydroxylase; Exonuclease; DNA-binding.  
 FT DOMAIN 1 318 5'-3' EXONUCLEASE.  
 FT DOMAIN 319 531 3'-5' EXONUCLEASE.  
 FT DOMAIN 532 908 POLYMERASE.  
 FT SEQUENCE 908 AA; 105503 MW; B58512AE80DCBD94 CRC64;

Query Match 3.3%; Score 130.5; DB 1; Length 908;  
 Best Local Similarity 19.3%; Pred. No. 5.4;  
 Matches 166; Conservative 103; Mismatches 264; Indels 327; Gaps 43;

13 LNTQRYLSKNSKRETVTLVGFMLFALVTSWGAQTVGLVESSTPAINPDSSSEYRW 72  
 26 LNTQ--GENVANF-----IGFF-----KTLFFIIKKRBEHLITTFDESEVPT 65  
 73 YGVESYVRGHPYKOR-----VAHDLRVNL----- 98  
 66 FRQKY-----PSYKATRDLPDDLPQIGWIKELGILKAKIPFEMEGYRADDLASFAK 121  
 99 EGSRSYQVCFN-----LKKAPLGSDSVYK--KMYKKHGGI-STKFEDYAMS 143  
 122 AAKNNITLYIISPKDLDTMSYVAILKIENNSFIEMONEVYTKKFGVNSQIKRYL-- 179  
 144 PRITDELNOKLRAVYNGHPONAN-----GIMEGLEPLN-----AIRVQEAIV 188  
 180 -AIVGDR-SDNIPGIGIGAKGANLLREFKTLGDIYSNLEIINKHRELLIKEKENAEL 237  
 189 YV-----SDNAPISND-----ESFKRESNLSVTSQSLMRQALKOLIDPNLATMPKQ 239

[illegible]

CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
DR	EMBL; AE001458; AAD05780.1; -
DR	HSPF; P07900; IYER.
DR	InterPro; IPR001404; -
DR	Pfam; PF00183; HSP90; 2.
DR	PRINTS; PR00775; HEATSHOCK90.
DR	PROSITE; PS00298; HSP90; 1.
KW	Chaperone; ATP-binding; Heat shock.
SO	SEQUENCE 621 AA; 71140 MW; 3CA62ABDD9796B69 CRC64;
Query Match	3.3%; Score 130; DB 1; Length 621;
Best Local Similarity	20.8%; Pred. No. 3.3; Indels 178; Gaps 27;
Matches 126; Conservative 78; Mismatches 225;	
QY	201 ESFKESES-----IVSISOLSLRQALKOLIDPMLATKMPQVDPOLSISESDEKG- 255
DB	28 ELFLAEELISNADALDKLNYLMLTDEKLT-----GLNTTPSHLSFDSQKTLTKDNGI 82
QY	256 --DKYN-----KGYONLLSGGLVETKPPTPGDPMPNPQDPTSVLRKXAI 300
DB	83 GMDKSDLIEHLGIATKSGTKSFLSALSG-----DKKDSALIGRGV 124
QY	301 GDYSKLBEATILQITGDVNVSFQARVFSSN-----DIGERIELSGTYTLPLNSPAGY- 354
DB	125 GYSAFPMVASKIVVQTKRVTSHQAAVWSVDGKGFELISCEVCEQGETITFLKEEDSHF 184
QY	355 -----SIAPITPKVAGKVTYITIDGKOLENPKKEIYEYSEAYNDPEEF- 400
DB	185 ASRWEIDSVYKKYSEHIFPPIFLTITDTKFEQ---EGDNKKYVEEKCQDINQASALMKM 241
QY	401 --SVLTQNYAKFYAKNKGSSQVYVCFNADLSPDSEDEGKTMTPTDTTGEVKKYTHI 458
DB	242 NKSLEKEDYKDFYQSPAHNDSEPLSYLHNK-----VGSLEYTTL 282
QY	459 -----AGDLFK-----YTKPRDTPD-----TELKHKKVLEKG-----YRE 492
DB	283 FYIPSKAPDLEFRVDYKSGVKLYVRVFTITDDDKELLSYLFKVGVIDSEDLPLNVSRE 342
QY	493 KQALIEYSLFTQLRANQQLAIYFTDSAEIDKDKIDYHFGSGMNSTLAVAKIIVE- 551
DB	343 ---LIQOKILIANRSASVKTI--LSELERLSKDN-KYTHKFEY-----PFGVYLEG 389
QY	552 -YAQDSNPQQLDLPFIFIPNNKYOSLIGTQHPREDLVDIIMEDKKEVIPIYTHNLTKLK 610
DB	390 LYGDENKEKLEL-----LRRYSNDKGWISLKEKKEKLKN-----QK 429
QY	611 TVTGLAGDFKDFHEIELKNKQK-----LLSQTV-----KTDKTNLEFGDKAT 656
DB	430 SIYUILLGEMLDLKASPLLEKAYQGYVDLLSDDEIDFVMPGVNEVDKT--PFDFA-- 484
QY	657 IMLKGESITLOGLE-----GYSYLKEDSDGKYKVVASQVANAATYKTSITSD 709
DB	485 ---SHSESLEKELGLAEIHDEVKDFKRLKAFE-ENLDELKIGVELSGHLTSAAVALIGDE 540
QY	710 TLAFFNN 716
DB	541 PNAAMN 547
RESULT 15	
FUSL_SCHPO	STANDARD: PRT: 1372 AA.
AC	Q10719;
DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	15-JUL-1998 (Rel. 36, Last annotation update)

